

Cálculo da Dimensão da Amostra em Estudos Clínicos: Princípios Metodológicos Básicos [119]

ANTÓNIO VAZ CARNEIRO

Centro de Estudos de Medicina Baseada na Evidência
Faculdade de Medicina de Lisboa, Lisboa, Portugal

Rev Port Cardiol 2003;22 (12):1513-1521

RESUMO

Os estudos clínicos, para serem válidos, devem ser metodologicamente rigorosos. O conceito de validade interna de um estudo é de crucial importância: um estudo é válido se os respectivos resultados representam uma estimativa não enviesada do resultado verdadeiro. Neste caso, a validade define-se como interna porque se refere ao grupo de doentes que estão a ser estudados e não necessariamente a outros (neste último caso seria a validade externa ou aplicabilidade). A validade interna da investigação clínica é determinada através do rigor do desenho do estudo, da recolha dos dados e das respectivas análises, podendo ser posta em causa por viéses (erros sistemáticos) ou pelo acaso (variação aleatória dos fenómenos sob estudo).

Independentemente do tipo de estudo (analítico, descritivo), as características da respectiva amostra de doentes é fundamental para a validade dos resultados. Os métodos de amostragem são fundamentais para a representatividade dos doentes do estudo em relação à população para a qual se deseja extrapolar os resultados conseguidos. Dentre as características fundamentais de uma amostra rigorosa encontra-se a sua correcta dimensão. Mesmo o estudo mais bem executado pode não conseguir responder capazmente à questão colocada, se a dimensão da respectiva amostra for muito pequena. Por outro lado, estudos com amostras de grandes dimensões são mais difíceis de concluir e muito mais dispendiosos. O objectivo do planeamento da dimensão correcta da amostra de um estudo

ABSTRACT

Estimating Sample Size in Clinical Studies: Basic Methodological Principles

In order to be valid, clinical studies must be methodologically rigorous. The internal validity of a study is of crucial importance: a study is valid if its results are an unbiased estimation of the true result. In this case, the validity is internal because it refers to the group of patients under study and not necessarily different ones (external validity or applicability). Internal validity in clinical research is achieved through rigorous design, data collection and appropriate analysis, and is threatened by bias (systematic errors) or chance (random variation of the phenomena under study).

Regardless of the type of study (analytic, descriptive, etc.), the characteristics of its sample are fundamental for the validity of the results. The sampling methods are crucial if the study patients are to be representative of the population to which one desires to extrapolate the results. One of the most fundamental characteristics of a sample is its size. Even the best executed study may fail to answer the research question if the sample size is too small. On the other hand, a study with too large a sample is harder to conduct and more costly. The goal of planning the sample size is to estimate the appropriate number of research subjects for the study. In this paper we will present and discuss the methodological principles underlying calculation of sample size: outcomes, type I and II error, alpha and beta, study power and variability.

consiste na estimativa do número apropriado de sujeitos para o estudo específico. Neste artigo iremos apresentar e discutir os princípios de base metodológica que servem para o cálculo da dimensão da amostra de um estudo: resultados (*outcomes*), erros tipo I e II, alfa e beta, poder do estudo e variabilidade.

Palavras-Chave

Hipóteses de investigação; Dimensão do efeito; Poder do estudo; Nível de significância; Erros tipo I e II; Alfa e beta; Dimensão da amostra; Validade interna; Cardiologia baseada na evidência

Key words

Research hypothesis; Effect size; Study power; Significance level; Type I and II error; Alpha and beta; Sample size; Internal validity; Evidence-based cardiology

INTRODUÇÃO

Os estudos clínicos, para serem válidos, devem ser metodologicamente rigorosos^(1, 2). O conceito de **validade interna** de um estudo é de crucial importância e varia ligeiramente conforme o tipo de questão clínica que se quer responder: no caso de estudos diagnósticos ou terapêuticos, por exemplo, eles são válidos se os respectivos resultados representam uma estimativa não enviesada do resultado verdadeiro.

Neste caso, a validade define-se como interna porque se refere ao grupo de doentes – amostra – que estão a ser estudados e não necessariamente a outros (neste último caso seria a validade externa ou aplicabilidade)⁽³⁾. A validade interna da investigação clínica é determinada através do rigor do desenho do estudo, da recolha dos dados e das respectivas análises, podendo ser posta em causa por viéses (erros sistemáticos) ou pelo acaso (variação aleatória dos fenómenos sob estudo)⁽⁴⁾.

Independentemente do tipo de estudo (analítico, descritivo), as características da respectiva amostra de doentes é fundamental para a validade dos resultados. Os métodos de amostragem são fundamentais para a representatividade dos doentes do estudo em relação à *população* para a qual se deseja extrapolar os resultados conseguidos. Dentre as características fundamentais de uma amostra rigorosa encontra-se a sua correcta dimensão: mesmo o estudo mais bem executado pode não conseguir responder capazmente à questão colocada, se a dimensão da respectiva amostra for muito pequena. Por outro lado, estudos com amostras de grandes dimensões são mais difíceis de concluir e muito mais dispendiosos. O objec-

INTRODUCTION

In order to be valid, clinical studies must be methodologically rigorous^(1, 2). The internal validity of a study is of crucial importance, and varies slightly depending on the type of clinical question being posed. Diagnostic or therapeutic trials, for example, are valid if their results are an unbiased estimation of the true result.

In this case, the validity is internal because it refers to the group of patients under study – the sample – and not necessarily different ones (which would be external validity or applicability)⁽³⁾. Internal validity in clinical research is achieved through rigorous design, data collection and appropriate analysis, and is threatened by bias (systematic errors) or chance (random variation of the phenomena under study)⁽⁴⁾.

Regardless of the type of study (analytic, descriptive, etc.), the characteristics of its sample are fundamental for the validity of the results. The sampling methods are crucial if the study patients are to be representative of the population to which one desires to extrapolate the results. One of the most fundamental characteristics of a sample is its size: even the best-executed study may fail to answer the research question if the sample size is too small. On the other hand, a study with too large a sample is harder to conduct and more costly. The goal of planning the sample size is to estimate the appropriate number of research subjects for the study⁽⁵⁾.

In this paper we will present and discuss the methodological principles underlying estimation of sample size. The discussion will not

tivo do planeamento da dimensão correcta da amostra de um estudo consiste na estimativa do número apropriado de sujeitos para o estudo específico⁽⁵⁾.

Neste artigo iremos apresentar e discutir os princípios de base metodológica que servem para o cálculo da dimensão da amostra de um estudo, mas não discutiremos nem as equações específicas, nem as instruções matemáticas utilizadas para cálculo do número final de doentes de uma amostra. Os leitores interessados podem consultar textos clássicos para esse efeito⁽⁶⁻¹²⁾.

CONSIDERAÇÕES GERAIS

A definição evolutiva da questão científica inicial que se pretende ver respondida por um determinado estudo constitui o que se designa como **hipótese de investigação**, que é uma versão final mais elaborada daquela e que inclui os elementos do estudo (desenho geral, amostragem, marcadores de doença, análises estatísticas, etc.). A colocação de uma hipótese é necessária na maior parte dos estudos observacionais e em todos os experimentais, que comparam, de algum modo, dois ou mais grupos de doentes^(13, 14).

A hipótese de investigação é muito importante num estudo, já que é nela que assenta a selecção dos testes de significância estatística⁽⁶⁾. Uma hipótese pode ser **simples**, se contém apenas um factor preditivo e uma variável de resultado (*outcome variable*): por ex., o *tabagismo associa-se a uma maior incidência de doença cerebrovascular*. Uma hipótese é **complexa** se contém mais do que um factor preditivo (*o tabagismo associado a hipertensão associa-se a uma maior incidência de doença cerebrovascular*), ou do que uma variável de resultado: por ex., *o tabagismo associa-se a uma maior incidência de doença cerebrovascular e de cancro do pulmão*. Uma hipótese é **específica** se não deixa dúvidas ou ambiguidade acerca dos sujeitos, variáveis e testes estatísticos a utilizar.

Em termos de testar significâncias estatísticas, a hipótese de investigação deve ser expressa em termos de diferenças esperadas entre os grupos de estudo. Classicamente, a **hipótese nula** afirma que não existe associação entre o factor preditivo e a variável na população ou que, por exemplo, não existe diferença entre os resultados de duas intervenções terapêuticas. A hipótese nula constitui a base

cover specific equations or the mathematical methods used to estimate the final number of patients in a sample. The interested reader can consult the standard texts in this area⁽⁶⁻¹²⁾.

GENERAL CONSIDERATIONS

The process of defining the initial clinical question for which an answer is sought in a particular trial leads to what is called the research hypothesis. This is a more detailed final version of the initial question, and covers the study elements (overall design, sampling, end-points, statistical analysis, etc.). It is necessary to put forward a hypothesis in most observational and all experimental studies that in one way or another compare two or more groups of patients^(13, 14).

The research hypothesis is a crucial part of a study, since it is the basis for selecting tests of statistical significance⁽⁶⁾. A hypothesis can be simple, if it contains only one predictive factor and one outcome variable (such as “Smoking is associated with a higher incidence of cerebrovascular disease”), or complex, if it has more than one predictive factor (“Smoking together with hypertension is associated with a higher incidence of cerebrovascular disease”) or more than one outcome variable (“Smoking is associated with a higher incidence of cerebrovascular disease and lung cancer”). A hypothesis is termed specific if there is no uncertainty or ambiguity concerning its subjects, variables and statistical tests.

With regard to testing statistical significance, the research hypothesis must be expressed in terms of the expected differences between the study groups. Classically, the null hypothesis states that there is no association between the predictive factor and the variable in the population; for example, that there is no difference between the results of two therapeutic interventions. The null hypothesis is the formal basis for testing statistical significance. By contrast, the alternative hypothesis is that there is an association between the predictive factor and the outcome variable. It is important to remember that the alternative hypothesis cannot be tested directly; it is accepted if the test of statistical significance rejects the null hypothesis⁽¹⁵⁾.

Finally, a one-sided hypothesis specifies the direction of the association between the predictive factor and the outcome variable,

formal para testar a significância estatística. Pelo contrário, a **hipótese alternativa** afirma a existência da associação entre o factor preditivo e a variável do resultado. É importante lembrar que a hipótese alternativa não pode ser testada directamente; ela é aceite se o teste estatístico de significância rejeita a hipóteses nula⁽¹⁵⁾.

Finalmente, uma hipótese unilateral (*one-sided*) especifica a direcção da associação entre o factor preditivo e a variável do resultado, por ex., *a hipertensão é mais frequente em doentes com doença cerebrovascular*. Uma hipótese bilateral (*two-sided*) afirma apenas que existe uma relação entre o factor preditivo e a variável do resultado, por ex., *em doentes com doença cerebrovascular, a tensão arterial é muito variável*. É mais fácil testar uma hipótese alternativa com um teste estatístico unilateral (*one-sided*) que com um bilateral (*two-sided*) – o erro tipo I é metade (ver adiante).

PRINCÍPIOS ESTATÍSTICOS

A importância da amostra em estudos clínicos – seja de que tipo for – já foi sublinhada anteriormente. Ela constitui uma das bases de qualquer estudo que se quer válido.

Para que um investigador possa determinar uma relação de causalidade entre um factor (por exemplo, uma intervenção terapêutica – *baixa do colesterol*) e um resultado (*diminuição da taxa de doença coronária*), ele deverá começar por assumir a hipótese nula de nenhuma associação. Depois, baseado nos dados finais coligidos na amostra do seu estudo, seleccionar os testes estatísticos ideais que lhe permitam determinar a existência de provas suficientes para rejeitar a hipótese nula. Ao fazê-lo, ele implicitamente decide a favor da hipótese alternativa (de que há uma associação entre a *baixa do colesterol* e a *diminuição da taxa de doença coronária*). O *standard* para estes testes é o chamado **nível de significância**⁽¹⁶⁾.

As bases sobre as quais assenta o cálculo da dimensão da amostra em estudos analíticos ou experimentais são várias, todas necessárias para um correcto resultado final, e incluem: os erros tipo I e II e as probabilidades de os cometer (designadas como alfa e beta), a dimensão do efeito, a variabilidade dos resultados e o poder do estudo. Na restante parte deste artigo iremos definir cada um destes factores.

such as “hypertension is more common in patients with cerebrovascular disease”. A two-sided hypothesis states only that there is a relation between the predictive factor and the outcome variable, such as “in patients with cerebrovascular disease, blood pressure is very variable”. It is easier to test an alternative hypothesis with a one-sided statistical test than with a two-sided one, as this halves the chances of a type I error (see below).

STATISTICAL PRINCIPLES

The importance of the sample size in any type of clinical trial has already been emphasized. It is one of the bases of the validity of any study.

An investigator who wishes to show a causal relation between a factor (such as a therapeutic intervention to lower cholesterol) and an outcome (a reduction in the rate of coronary artery disease) should begin by assuming the null hypothesis of no association. Then, based on the final data collected in the study sample, the most suitable statistical tests should be selected that will determine whether there is sufficient evidence to reject the null hypothesis. Such rejection implicitly decides in favor of the alternative hypothesis, that there is an association between lowering cholesterol and a reduction in the rate of coronary artery disease. The standard for these tests is the level of significance⁽¹⁶⁾.

There are several aspects relevant to estimating the correct sample size in analytic or experimental trials, all of which must be taken into consideration. They include type I and II error and the probability of committing them (termed alpha and beta), effect size, variability in outcomes, and study power. The rest of this article will discuss each of these factors.

Type I and II error, alpha (α) and beta (β)

As already stated, at the end of the trial the investigator will use statistical tests to attempt to reject the null hypothesis in favor of the alternative hypothesis. Assuming that there was no bias that altered the results, and depending on whether the null hypothesis was true or false in the general population, there are four possibilities (*Table I*):

Erros tipo I e II, alfa (α) e beta (β)

Como já foi dito, no final do estudo o investigador vai lançar mão de testes estatísticos na tentativa de rejeitar a hipótese nula e favorecer a alternativa. Partindo do princípio que não existiram viéses que alterem sistematicamente os resultados e dependendo da hipótese nula ser verdadeira ou falsa na população em geral, são possíveis quatro hipóteses (*Quadro I*).

Quadro I

Verdade na população vs. resultados na amostra do estudo

Resultados na amostra do estudo	Verdade na população	
	Existe associação entre o factor e o resultado	Não existe associação entre o factor e o resultado
Rejeita a hipótese nula	Correcto	Erro tipo I
Não rejeita a hipótese nula	Erro tipo II	Correcto

- O erro tipo I (falso-positivo) ocorre se um investigador rejeita a hipótese nula, quando ela é verdadeira na população;
- O erro tipo II (falso-negativo) ocorre se um investigador não rejeita a hipótese nula, quando ela é falsa na população.

Os erros tipo I e II não podem ser evitados por completo, mas pode diminuir-se a sua probabilidade aumentando a dimensão da amostra do estudo, por exemplo (porque se aproxima a dimensão da amostra da dimensão da população da qual aquela emanou, reduzindo as diferenças eventualmente existentes entre elas)⁽⁶⁾.

A probabilidade (previamente definida) de um investigador cometer um erro tipo I (rejeitar a hipótese nula quando ela é verdadeira) designa-se por **alfa** (α) ou, como também é conhecido, por **nível de significância estatística**. Por exemplo, se num estudo relacionando os efeitos do exercício com os níveis de HDL o alfa é de 0,05, então o investigador define 5% como a chance máxima de incorrectamente rejeitar a hipótese nula (inferindo que o exercício e os níveis de HDL estão associados na população, quando de facto não estão). O alfa tem relação com o valor de *P*, determinado pelos testes estatísticos, já que a hipótese nula é rejeitada se o valor de *P* é menor que o alfa^(17, 18).

Por outro lado, a probabilidade de um investigador cometer um erro tipo II (não rejeitar a hipótese nula quando ela é de facto falsa) designa-se por **beta** (β). Idealmente, o alfa e o beta deveriam ser iguais a zero, mas isto é im-

Table I

Truth in the population vs. results in the study sample

Results in the study sample	Truth in the population	
	There is an association between the factor and the outcome	There is no association between the factor and the outcome
Reject the null hypothesis	Correct	Type I error
Do not reject the null hypothesis	Type II error	Correct

- A type I error (false positive) occurs if the investigator rejects the null hypothesis when it is in fact true in the population;
- A type II error (false negative) occurs if the investigator does not reject the null hypothesis when it is in fact false in the population.

Type I and II errors cannot be avoided completely, but they can be made less likely by such strategies as increasing the study's sample size, as this will bring the sample size nearer to that of the population from which the sample was taken and thus reduce any differences between them⁽⁶⁾.

The predefined likelihood of an investigator committing a type I error (rejecting the null hypothesis when it is true) is termed alpha (α), and is also known as the level of statistical significance. For instance, in a study relating the effects of exercise to HDL levels, if alpha is 0.05, then the investigator is setting 5% as the maximum chance of wrongly rejecting the null hypothesis (i.e. inferring that exercise and HDL levels are associated in the population, when in fact they are not). Alpha is related to the *p* value, as determined by statistical tests, since the null hypothesis is rejected if *p* is less than alpha^(17, 18).

Conversely, the probability of an investigator committing a type II error (failing to reject the null hypothesis when it is in fact false) is termed beta (β). Ideally, alpha and beta should both be equal to zero, but in practice this is impossible. Errors can be made less likely by increasing sample size, using continuous variables or more precise or paired measures, or by using different-sized groups or more common outcomes^(7, 19).

If it is particularly important to avoid type I errors, such as when testing a high-risk treatment, alpha should be set lower. On the other hand, if it is more important to avoid type II errors (false positives), it is beta that should be set to a lower level.

possível. Podemos diminuir a possibilidade da sua ocorrência aumentando a dimensão da amostra, utilizando variáveis contínuas, medições mais precisas ou emparelhadas, utilizando grupos de dimensões desiguais ou utilizando resultados (*outcomes*) mais comuns^(7, 19).

Se para o investigador é particularmente importante evitar um erro tipo I (por exemplo quando quer testar a eficácia de uma intervenção terapêutica com riscos elevados), então ele deve colocar o alfa a níveis mais baixos. Pelo contrário, se quiser evitar o erro tipo II (falso-negativo), então é o beta que deverá ser colocado a níveis baixos.

Dimensão do efeito

A possibilidade de um estudo detectar uma associação entre um factor preditivo e um resultado depende da dimensão do efeito esperado: se este for de grandes dimensões, é de mais fácil detecção. O problema surge quando o investigador não tem uma estimativa prévia da magnitude da associação. Neste caso, não haverá outra hipótese senão ele decidir *a priori* a dimensão que gostaria de detectar ou, alternativamente, que considere clinicamente significativa: por exemplo, uma diminuição média de 30 mmHg da pressão sistólica num estudo sobre a eficácia de um medicamento anti-hipertensivo. Esta medida designa-se por **dimensão do efeito**⁽¹¹⁾.

Este passo é de enorme importância no desenho do estudo e pode, se for mal calculado, comprometer ou anular os resultados. Existem três hipóteses de o investigador minimizar esta incerteza: a primeira é basear-se em dados publicados que, de qualquer maneira, possa fornecer estimativas realistas; a segunda, é aceitar a menor dimensão que seja clinicamente significativa (no exemplo anterior, 20 mmHg); a terceira pode obrigar a um estudo piloto, especialmente quando os dados são pura e simplesmente inexistentes.

Uma última precaução: para efeitos de cálculo da dimensão amostral, a medida de associação deve basear-se no resultado (*outcome*) principal. As dimensões dos efeitos para os resultados secundários podem ser calculadas depois a partir dessa amostra.

Poder do estudo

Já vimos anteriormente a definição dos erros tipo I e II. Importa agora definir o conceito de **poder** do estudo. Este define-se como o inverso de β , isto é, $1 - \beta$.

Effect size

The ability of a trial to detect an association between a predictive factor and an outcome depends on the size of the expected effect. The larger the effect, the easier it is to detect. Problems arise when the investigator has no previous estimate of the magnitude of the association. In this case, there is no alternative to deciding *a priori* what size effect one would like to detect or would consider clinically significant, such as a mean fall of 30 mmHg in systolic blood pressure in a trial on the efficacy of an antihypertensive drug. This is known as the effect size⁽¹¹⁾.

Calculating the effect size is a very important step in designing the study, as a miscalculation can compromise or even invalidate the results. There are three ways in which the investigator can minimize the uncertainty. The first is to base the calculation on published data which can help arrive at a realistic estimate; secondly, to accept the smallest clinically significant effect size (which could be 20 mmHg in the example given); thirdly to carry out a pilot study if necessary, especially when there is simply no data available.

As a final precaution, when calculating sample size, the measure of association should be based on the primary outcome. Effect sizes for secondary outcomes can then be calculated on the basis of this sample.

Study power

Type I and II errors have now been described; we now need to define the concept of the study's power. This is calculated as the inverse of β , that is, $1 - \beta$.

The power of a study is the probability of rejecting the null hypothesis in that sample if the real effect in the population is equal to the effect size. If beta is predefined as 0.20, this means that the investigator is willing to accept a 20% possibility of failure to detect a possible association. The study power is 0.80, meaning that the study has an 80% chance of detecting an association with the specified effect. Most studies set alpha at 0.05 (ranging between 0.01 and 0.10) and beta at 0.20 (ranging between 0.05 and 0.20).

In order to summarize all these factors and their interaction, Table II shows the variables used to calculate the power of a study using a paired Student's t test and their effect on sample size⁽⁹⁾.

O poder de um estudo é a probabilidade de rejeitar a hipótese nula daquela amostra, se o efeito real na população é igual à dimensão do efeito. Se o β é pré-definido como 0,20, isto quer dizer que o investigador está disposto a aceitar 20 % de hipótese de falhar a detecção de uma eventual associação. O poder do estudo é de 0,80, isto é, o estudo tem uma chance de 80 % de detectar uma associação com o efeito determinado. A maior parte dos estudos coloca o alfa a 0,05 (podendo variar entre 0,01 e 0,10) e o beta a 0,20 (podendo variar entre 0,05 e 0,20).

Como resumo de todos estes factores e suas interacções, estão descritas no *Quadro II* as variáveis incluídas no cálculo do poder de um estudo utilizando um teste de *t* emparelhado e o seu efeito na dimensão da amostra⁽⁹⁾.

Quadro II

Variáveis incluídas no cálculo do poder de um estudo utilizando um teste de *t* emparelhado e o seu efeito na dimensão da amostra*

Variável	δ	σ	α	$1-\beta$	n
Teste bilateral**	5	20	0,05	0,8	127
Teste unilateral	5	20	0,05	0,8	100
Aumento de δ	10	20	0,05	0,8	25
Aumento de σ	5	25	0,05	0,8	155
Diminuição de α	5	20	0,01	0,8	160
Diminuição de β	5	20	0,05	0,9	138

* Os valores a negrito foram variados a partir da primeira linha, para demonstrar como as alterações em cada variável afectam a dimensão da amostra.

** Os testes bilaterais são mais conservadores que os unilaterais, porque, para se obter significância estatística são necessárias amostras de maiores dimensões.

- δ = (delta) dimensão da diferença ou alteração a ser detectada – idealmente, a diferença mais pequena julgada clinicamente significativa pelo investigador.
- σ = (sigma) o desvio padrão; é o aumento da variabilidade da diferença entre as observações emparelhadas.
- α = (alfa) o limiar abaixo do qual se define a significância estatística (definido pelo investigador).
- $1-\beta$ = o poder estatístico do estudo (definido pelo investigador).
- n = dimensão da amostra.

Adaptado de⁽⁹⁾

Para finalizar: as fórmulas utilizadas no cálculo da dimensão amostral são baseadas em conceitos matemáticos diferentes, conforme os testes estatísticos seleccionados. Estes devem ser escolhidos antes do início do estudo e dependem do tipo de factor preditivo e da variável incluída. No *Quadro III* estão incluídos alguns testes estatísticos usados neste contexto⁽¹¹⁾.

Table II

Variables used to calculate the power of a study using a paired Student's *t* test and their effect on sample size*

Variable	δ	σ	α	$1-\beta$	n
Two-sided test**	5	20	0.05	0.8	127
One-sided test	5	20	0.05	0.8	100
Increased δ	10	20	0.05	0.8	25
Increased σ	5	25	0.05	0.8	155
Decreased α	5	20	0.01	0.8	160
Decreased β	5	20	0.05	0.9	138

* The values in bold have been varied from those in the first line to demonstrate how changes in each variable affect the sample size.

** Two-sided tests are more conservative than one-sided tests, since larger sample sizes are required to reach statistical significance.

- δ = (delta) = size of the difference or change to be detected – ideally the smallest difference considered clinically significant by the investigator.
- σ = (sigma) standard deviation; the increase in variability of the difference between paired observations.
- α = (alpha) = the limit below which statistical significance is deemed by the investigator to have been reached.
- $1-\beta$ = the statistical power of the study (as defined by the investigator).
- n = sample size.

Adapted from⁽⁹⁾

A final note: the formulas used to calculate the sample size are based on different mathematical concepts, depending on the statistical tests selected. These tests should be chosen before the beginning of the trial and depend on the type of predictive factor and the variable included. *Table III* shows some statistical tests used for this purpose⁽¹¹⁾.

Table III

Statistical tests used to determine sample size

	Variable (outcome)	
Predictive factor	Dichotomous	Continuous
Dichotomous	Chi-square	Student's <i>t</i> test
Continuous	Student's <i>t</i> test	Correlation coefficient

CONCLUSIONS

For both scientific and ethical reasons, sample size in clinical trials must be carefully planned if the results are to be credible⁽¹⁴⁾. Ideally, the trial should be large enough to have a high probability of detecting a clinically important difference of a certain magnitude, if such a difference really exists, and showing it to be statistically significant⁽²⁰⁾.

Studies, whether analytic or experimental (purely descriptive studies do not need hypotheses), require a hypothesis that specifies the expected association between the predictive factors and the outcomes, in order for tests of

Quadro III

Testes estatísticos para determinação da dimensão amostral

Factor preditivo	Variável (outcome)	
	Dicotómica	Contínua
Dicotómica	Chi-quadrado	Teste <i>t</i>
Contínua	Teste <i>t</i>	Coefficiente de correlação

CONCLUSÕES

A dimensão da amostra em estudos clínicos necessita, por razões científicas e éticas, de ser bem planeada, se se quiser que os resultados sejam credíveis⁽¹⁴⁾. Idealmente, um estudo deverá ter uma dimensão suficientemente importante para poder ter uma alta probabilidade de detectar, como estatisticamente significativa, uma diferença clinicamente importante com uma determinada dimensão, se essa diferença realmente existir⁽²⁰⁾.

Os estudos – quer sejam analíticos, quer experimentais (os puramente descritivos não necessitam de hipóteses) – necessitam da colocação de uma hipótese que especifique (para aplicação ulterior de testes de significância) a associação esperada entre os factores preditivos e os resultados. A hipótese nula – que é a base formal para testar a significância estatística – afirma que não existe associação entre o factor preditivo e a variável do resultado na população ou que, por exemplo, não existe diferença entre os resultados de duas intervenções terapêuticas. A hipótese alternativa – que poderá ser uni ou bilateral – afirma a existência da associação entre o factor preditivo e a variável do resultado medido.

Os elementos para o cálculo correcto da dimensão da amostra – baseada no resultado principal que se quer medir – são: 1) os resultados (*outcomes*) em cada grupo, na base dos quais se define o que são as diferenças clinicamente significativas; 2) o erro tipo I (alfa); 3) o poder do estudo (o nível de erro tipo II ou beta) e 4) (no caso de variáveis contínuas) o desvio padrão das medidas – a variabilidade.

statistical significance to be applied subsequently. The null hypothesis, the formal basis for testing statistical significance, states that there is no association between the predictive factor and the outcome variable in the population; that, for example, there is no difference between the outcomes of two therapeutic interventions. The alternative hypothesis (which can be one- or two-sided) states that there is in fact an association between the predictive factor and the measured outcome variable.

The elements required for correct calculation of the sample size – based on the primary outcome that is to be measured – are: 1) the outcomes in each group, on the basis of which clinically significant differences are determined; 2) type I error (alpha); 3) the study power (the level of type II error, or beta); and 4) in the case of continuous variables, the standard deviation of the measures, or variability.

Pedido de separatas para:

Address for reprints:

ANTÓNIO VAZ CARNEIRO

Centro de Estudos de Medicina

Baseada na Evidência

Faculdade de Medicina de Lisboa

Av. Prof. Egas Moniz

1649-035 LISBOA Portugal

BIBLIOGRAFIA / REFERENCES

1. Carneiro AV. A Medicina Baseada na Evidência. Um novo paradigma para a prática médica. *Acta Méd Port* 1998;11:717-37.
2. Carneiro AV. Um instrumento para a mudança da prática médica: a "Medicina Baseada na Evidência". *Medicina Int* 1998;5:133-8.
3. Guyatt GH, Rennie D, editors. *User's Guides to the Medical Literature. A Manual for Evidence-Based Clinical Practice*. Chicago: JAMA & Archives Journals AMA, 2001.
4. *Clinical epidemiology*. Baltimore: Williams & Wilkins, 1996.
5. Campbell MJ, Julious SA, Altman DG. Estimating sample sizes for binary, ordered categorical, and continuous outcomes in two group comparisons. *BMJ* 1995;311:1145-8.
6. *Designing clinical research*. Philadelphia: Lippincott Williams and Wilkins, 2001.
7. *An introduction to medical statistics*. Oxford: Oxford University Press, 2000.
8. *Clinical epidemiology & evidence-based medicine*. Thousand Oaks: Sage Publications, 2001.
9. *How to report statistics in medicine*. Philadelphia: ACP, 1997.
10. *Statistical methods in diagnostic medicine*. New York: Wiley, 2002.
11. *Medical statistics*. Malden: Blackwell Science, 2003.
12. *Basic & clinical biostatistics*. New York: Lange Medical Books/McGraw-Hill, 2001.
13. Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Int Med* 2003;138:644-50.
14. Chan KBY, Man-Son-Ing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *CMAJ* 2001;165:1197-202.
15. Kerry SM, Bland JM. Sample size in cluster randomisation. *BMJ* 1998;316:549.
16. *Fundamentals of clinical trials*. New York: Springer, 1998.
17. Goodman SN. Toward evidence-based medical statistics. 1. The P value fallacy. *Ann Int Med* 1999;130:995-1004.
18. Goodman SN. Toward evidence-based medical statistics. 2. The Bayes factor. *Ann Int Med* 1999;130:1005-13.
19. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002;359:781-5.
20. Sackett DL. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *CMAJ* 2001;165:1226-37.

Hospital Fernando Fonseca

Serviço de Cardiologia

IV INTERNATIONAL SYMPOSIUM

Thrombosis and Thromboembolism

Penha-Longa Golf Resort
SINTRA – PORTUGAL

October, 1 and 2, 2004